

# BranchHull: Convex bilinear inversion from the entrywise product of signals with known signs

Alireza Aghasi\*, Ali Ahmed† and Paul Hand‡

February 15, 2017

## Abstract

We consider the bilinear inverse problem of recovering two vectors,  $\mathbf{x}$  and  $\mathbf{w}$ , in  $\mathbb{R}^L$  from their entrywise product. For the case where the vectors have known signs and belong to known subspaces, we introduce the convex program BranchHull, which is posed in the natural parameter space and does not require an approximate solution or initialization in order to be stated or solved. Under the structural assumptions that  $\mathbf{x}$  and  $\mathbf{w}$  are the members of known  $K$  and  $N$  dimensional random subspaces, we prove that BranchHull recovers  $\mathbf{x}$  and  $\mathbf{w}$  up to the inherent scaling ambiguity with high probability whenever  $L \gtrsim K + N$ . This program is motivated by applications in blind deconvolution and self-calibration.

## 1 Introduction

This paper considers a bilinear inverse problem (BIP): recover vectors  $\mathbf{x}$  and  $\mathbf{w}$  from the observation  $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{w})$ , where  $\mathcal{A}$  is a bilinear operator. BIPs have been extensively studied in signal processing and data science literature, and comprise of fundamental problems such as blind deconvolution/demodulation [3, 22, 14, 1], phase retrieval [8], dictionary learning [24], matrix factorization [13, 15], and self-calibration [17]. Optimization problems involving bilinear terms and constraints also arise in other contexts, such as blending problems in chemical engineering [6].

---

\*Department of Mathematical Sciences, IBM TJ Watson Research Center, Yorktown Heights, NY

†School of Electrical and Computer Engineering at Information Technology University, Lahore, Pakistan

‡Department of Computational and Applied Mathematics, Rice University, TX

A significant challenge of BIPs is the ambiguity of solutions. For example, if  $(\mathbf{x}^{\mathfrak{h}}, \mathbf{w}^{\mathfrak{h}})$  is a solution to a BIP, then so is  $(c\mathbf{x}^{\mathfrak{h}}, c^{-1}\mathbf{w}^{\mathfrak{h}})$  for any nonzero  $c \in \mathbb{R}$ . Other ambiguities may also arise, including the shift ambiguity in blind deconvolution, the permutation ambiguity in dictionary learning, and the ambiguity up to multiplication by an invertible matrix in matrix factorization. These ambiguities are challenging because they cause the set of solutions to be nonconvex.

We will consider the fundamental bilinear inverse problem of recovering two  $L$  dimensional vectors  $\mathbf{w}$  and  $\mathbf{x}$  from the observations  $\mathbf{y} = \mathbf{w} \circ \mathbf{x}$ , where  $\circ$  denotes the entry-wise product of vectors. This is immediately recognized as the calibration problem, where one is only able to measure a signal  $\mathbf{x}$  modulo unknown multiplicative gains  $\mathbf{w}$ . A self-calibration algorithm aims to figure out the gains  $\mathbf{w}$  and the signal  $\mathbf{x}$  jointly from  $\mathbf{y}$ . The circular convolution also becomes pointwise multiplication in the Fourier domain, allowing us to reduce the important blind deconvolution problem in signal processing and wireless communications to a complex case of the above bilinear form.

In addition to the challenges of general BIPs, the BIP above is difficult because the solutions are nonunique without further structural assumptions. For example  $(\mathbf{w}^{\mathfrak{h}}, \mathbf{x}^{\mathfrak{h}})$  and  $(1, \mathbf{w}^{\mathfrak{h}} \circ \mathbf{x}^{\mathfrak{h}})$  are both consistent with the entrywise products  $\mathbf{y} = \mathbf{w}^{\mathfrak{h}} \circ \mathbf{x}^{\mathfrak{h}}$ . While multiple structural assumptions are reasonable, we will consider the case where  $\mathbf{w}^{\mathfrak{h}}$  and  $\mathbf{x}^{\mathfrak{h}}$  belong to known subspaces  $\mathbf{B}$  and  $\mathbf{C}$ , as in [3]. In addition, we also require  $\mathbf{w}^{\mathfrak{h}}$  and  $\mathbf{x}^{\mathfrak{h}}$  to be real and of known signs. The method can be extended to complex vectors in the case of known complex phases.

The known sign information in the real case is justified in imaging applications, where we want to recover image pixels (always non-negative) from occlusions caused by some unknown multiplicative masks. A stylized application of this setup arises in the wireless communications. A source encodes a message as a series of positive magnitude shifts on tones at frequencies  $f_1, f_2, \dots, f_L$ . These real valued and positive  $\mathbf{x} = [x(f_1), x(f_2), \dots, x(f_L)]^{\top}$  are transmitted over a linear-time invariant channel, where  $x(f_{\ell})$  are weighted by the frequency response of the channel  $w(f_{\ell})$  (in general complex valued), and then in the ideal noiseless case, the receiver ends up observing  $y(f_{\ell}) = x(f_{\ell}) \cdot w(f_{\ell})$ . The real part of the complex-valued measurements  $\text{Re}\{y(f_{\ell})\} = x(f_{\ell}) \cdot \text{Re}\{w(f_{\ell})\}$  are simply the pointwise product of two unknown real numbers with known signs. In addition, in this application, the vectors  $\mathbf{x}$ , and  $\mathbf{w}$  naturally live in low-dimensional subspaces; for details, see [3, 2].

The bilinear inverse problem we consider is:

$$\begin{aligned}
&\text{Let: } \mathbf{w}^{\mathfrak{h}} \in \mathbf{B} \subset \mathbb{R}^L, \mathbf{x}^{\mathfrak{h}} \in \mathbf{C} \subset \mathbb{R}^L, \mathbf{y} = \mathbf{w}^{\mathfrak{h}} \circ \mathbf{x}^{\mathfrak{h}}, \mathbf{s} = \text{sign}(\mathbf{w}^{\mathfrak{h}}) \\
&\text{Given: } \mathbf{y}, \mathbf{s}, \mathbf{B}, \mathbf{C} \\
&\text{Find: } \mathbf{w}^{\mathfrak{h}}, \mathbf{x}^{\mathfrak{h}} \text{ up to the scaling ambiguity}
\end{aligned} \tag{1}$$

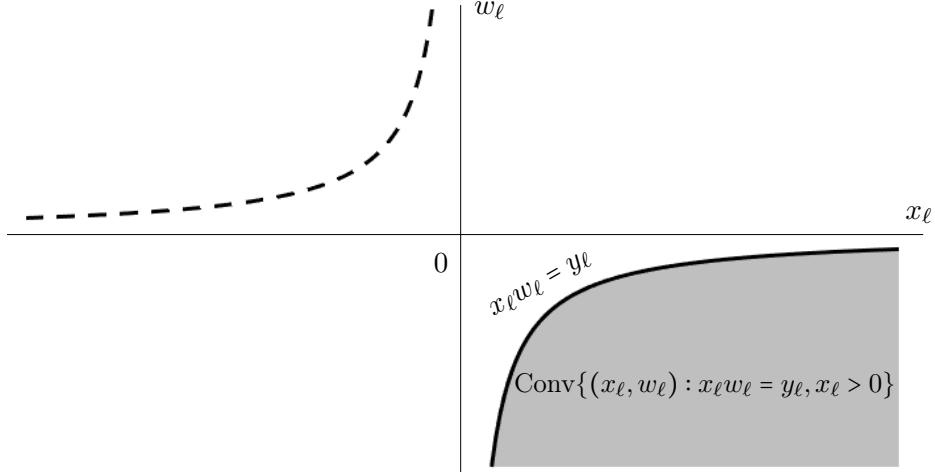
One way to solve the BIP above<sup>1</sup> is by lifting. More specifically, the bilinear inverse problem can be recast as a linear matrix recovery problem with the structural constraint that the recovered matrix is rank one. With  $w_\ell^{\mathfrak{h}} = \mathbf{b}_\ell^\top \mathbf{h}^{\mathfrak{h}}$  and  $x_\ell^{\mathfrak{h}} = \mathbf{c}_\ell^\top \mathbf{m}^{\mathfrak{h}}$  for  $\ell = 1, \dots, L$ , the underlying linear operator is given by  $y_\ell = w_\ell^{\mathfrak{h}} x_\ell^{\mathfrak{h}} = \mathbf{b}_\ell^\top \mathbf{h}^{\mathfrak{h}} \mathbf{m}^{\mathfrak{h}^\top} \mathbf{c}_\ell = \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \mathbf{h}^{\mathfrak{h}} \mathbf{m}^{\mathfrak{h}^\top} \rangle = \mathcal{A}_\ell(\mathbf{h}^{\mathfrak{h}} \mathbf{m}^{\mathfrak{h}^\top})$ , and the formal recovery framework is find the  $\mathbf{X}$  of minimal rank that is consistent with  $\mathcal{A}(\mathbf{X}) = \mathbf{y}$ . By relaxing the rank objective to the nuclear norm of  $\mathbf{X}$ , this optimization problem becomes a semidefinite program. The results in [3], which apply to the complex case, show that when  $\mathbf{b}_\ell$  and  $\mathbf{c}_\ell$  are Fourier and Gaussian vectors, respectively, this semidefinite program succeeds in recovering the rank-1 matrix  $\mathbf{h}^{\mathfrak{h}} \mathbf{m}^{\mathfrak{h}^\top}$  with high probability, whenever  $K + N \lesssim L / \log^3 L$ . Unfortunately, directly optimizing a lifted problem is prohibitively computationally expensive, as the lifted semidefinite program is posed on a space of dimensionality  $K \times N$ , which is much larger than the  $K + N$  dimensionality of the natural parameter space.

To address the intractability of lifted methods, a recent theme of research has been to solve quadratic and bilinear recovery problems in the natural parameter space using alternating minimization and gradient descent algorithms [19, 23]. These algorithms include the Wirtinger Flow (WF) and its variants for phase retrieval [5, 7, 26]. A Wirtinger gradient descent method was recently introduced for blind deconvolution in [16]. In the case that  $\mathbf{b}_\ell$  are deterministic complex matrices that satisfy an incoherence property and that  $\mathbf{c}_\ell$  are Gaussian vectors, this nonconvex method succeeds at recovering  $\mathbf{h}^{\mathfrak{h}}$  and  $\mathbf{m}^{\mathfrak{h}}$  up to the scale ambiguity with high probability when  $K + N \lesssim L / \log^2 L$ . While WF based methods enjoy rigorous recovery guarantees under optimal or nearly optimal sample complexity with suitable measurement models, the proofs of these results are long and technical. Also, because of the nonconvexity of the problem, the convergence of a gradient descent to the global minimum usually relies on an appropriate initialization [25, 7, 16].

Recently, a convex formulation for phase retrieval was found for the natural parameter space. This method, called PhaseMax, is a linear program

---

<sup>1</sup>As stated, this approach ignores the sign information.



**Figure 1:** Given the bilinear measurement  $x_\ell w_\ell = y_\ell$ , the point  $(x_\ell, w_\ell)$  is on a two-branch hyperbola, as depicted by the dashed and solid lines. Further information on the sign of  $w_\ell$  identifies which branch of the hyperbola the point is on (the solid line). The convex formulation in this paper replaces the relevant branch of the hyperbola with its convex hull (the shaded region).

and was independently discovered by [4] and [9]. PhaseMax enjoys a rigorous recovery guarantee under a random data model. Existing recovery proofs are based on statistical learning theory [4], geometric probability [9], and elementary probabilistic concentration arguments [11]. As with Wirtinger Flow, successful recovery of PhaseMax with optimal sample complexity has been proven when an appropriate initialization is known. Unlike Wirtinger Flow, the initialization is used in PhaseMax’s objective function, as opposed to its algorithmic implementation.

In this paper, we introduce a convex formulation in the natural parameter space for the bilinear inverse problem of recovering two real vectors from their entrywise product, provided that the vectors live in known subspaces and have known signs. The convex formulation is called BranchHull and is based on the following idea: The bilinear measurements  $x_\ell w_\ell = y_\ell$  establish that  $(x_\ell, w_\ell)$  is on one of two branches of a hyperbola in  $\mathbb{R}^2$ . Information on  $\text{sign}(w_\ell)$  identifies the appropriate branch. The convex formulation is then formed by relaxing this nonconvex branch of a hyperbola to its convex hull, as shown in Figure 1. This formulation does not require an approximate solution or initialization in order to be stated or solved. In the case where the two vectors live in random subspaces of  $\mathbb{R}^L$  with dimensions  $K$  and  $N$ , we establish that the two vectors can be recovered up to the inherent scaling

ambiguity with high probability for  $K + N \lesssim L$ . Further, we provide an explicit lower bound on the recovery probability that is nonzero when  $L > 4(K + N) - 3$ .

### 1.1 Problem Formulation

We consider the bilinear inverse problem of recovering two vectors from their entrywise product. That is, let  $\mathbf{w}^{\mathfrak{h}}, \mathbf{x}^{\mathfrak{h}} \in \mathbb{R}^L$ , and let  $\mathbf{y} = \mathbf{w}^{\mathfrak{h}} \circ \mathbf{x}^{\mathfrak{h}}$ . From  $\mathbf{y}$ , we attempt to find  $\mathbf{w}^{\mathfrak{h}}$  and  $\mathbf{x}^{\mathfrak{h}}$  up to the scaling ambiguity  $(c\mathbf{w}^{\mathfrak{h}}, \frac{1}{c}\mathbf{x}^{\mathfrak{h}})$ . To make the problem well posed, we consider the case where  $\mathbf{w}^{\mathfrak{h}}$  and  $\mathbf{x}^{\mathfrak{h}}$  belong to known subspaces  $\mathbf{B}$  and  $\mathbf{C}$  of  $\mathbb{R}^L$ . We further consider the case where the signs of the entries of  $\mathbf{w}^{\mathfrak{h}}$ , and hence those of  $\mathbf{x}^{\mathfrak{h}}$ , are known. Let  $\mathbf{s} = \text{sign}(\mathbf{w}^{\mathfrak{h}})$ . This bilinear inversion problem is stated in (1).

Ideally, we could resolve the scaling ambiguity and find  $(\mathbf{w}^{\mathfrak{h}}, \mathbf{x}^{\mathfrak{h}})$  such that  $\|\mathbf{w}^{\mathfrak{h}}\|_2 = \|\mathbf{x}^{\mathfrak{h}}\|_2$  by solving the following program:

$$\begin{aligned} \underset{\mathbf{w} \in \mathbf{B}, \mathbf{x} \in \mathbf{C}}{\text{minimize}} \quad & \|\mathbf{w}\|_2^2 + \|\mathbf{x}\|_2^2 \text{ subject to } w_\ell x_\ell = y_\ell \\ & s_\ell w_\ell \geq 0, \ell = 1, \dots, L. \end{aligned}$$

This program is nonconvex, but it admits the following convex relaxation:

$$\begin{aligned} \underset{\mathbf{w} \in \mathbf{B}, \mathbf{x} \in \mathbf{C}}{\text{minimize}} \quad & \|\mathbf{w}\|_2^2 + \|\mathbf{x}\|_2^2 \text{ subject to } \text{sign}(y_\ell) w_\ell x_\ell \geq |y_\ell| \\ & s_\ell w_\ell \geq 0, \ell = 1, \dots, L. \end{aligned}$$

Note that for fixed  $\ell$ , the feasible set  $\{(w_\ell, x_\ell) \mid \text{sign}(y_\ell) w_\ell x_\ell \geq |y_\ell|, s_\ell w_\ell \geq 0\}$  is the convex hull of  $\{(w_\ell, x_\ell) \mid w_\ell x_\ell = y_\ell, s_\ell w_\ell \geq 0\}$ .

We consider this problem when written in the natural parameter space. Let  $\mathbf{B} \in \mathbb{R}^{L \times K}$  be a matrix that spans the  $K$  dimensional subspace  $\mathbf{B}$ . Similarly, let  $\mathbf{C} \in \mathbb{R}^{L \times N}$  be a matrix that spans the  $N$  dimensional subspace  $\mathbf{C}$ . Let  $(\mathbf{h}^{\mathfrak{h}}, \mathbf{m}^{\mathfrak{h}}) \in \mathbb{R}^K \times \mathbb{R}^N$ . Let  $\mathbf{w}^{\mathfrak{h}} = \mathbf{B}\mathbf{h}^{\mathfrak{h}}$  and  $\mathbf{x}^{\mathfrak{h}} = \mathbf{C}\mathbf{m}^{\mathfrak{h}}$ . We can write  $w_\ell = \mathbf{b}_\ell^\top \mathbf{h}$ ,  $x_\ell = \mathbf{c}_\ell^\top \mathbf{m}$ , and  $y_\ell = \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \mathbf{h}^{\mathfrak{h}} \mathbf{m}^{\mathfrak{h}^\top} \rangle$ , where  $\mathbf{b}_\ell^\top$  is the  $\ell$ th row of  $\mathbf{B}$  and  $\mathbf{c}_\ell^\top$  is the  $\ell$ th row of  $\mathbf{C}$ . The recovery task is now to find  $(\mathbf{h}^{\mathfrak{h}}, \mathbf{m}^{\mathfrak{h}})$  by the convex program called BranchHull

$$\begin{aligned} \underset{\mathbf{h} \in \mathbb{R}^K, \mathbf{m} \in \mathbb{R}^N}{\text{minimize}} \quad & \|\mathbf{h}\|_2^2 + \|\mathbf{m}\|_2^2 \text{ subject to } \text{sign}(y_\ell) \mathbf{b}_\ell^\top \mathbf{h} \cdot \mathbf{c}_\ell^\top \mathbf{m} \geq |y_\ell| \quad (\text{BH}) \\ & s_\ell \cdot \mathbf{b}_\ell^\top \mathbf{h} \geq 0, \ell = 1, \dots, L. \end{aligned}$$

This program is convex because for any fixed  $\ell$ , the points consistent with both the first and second constraints is a convex set. This program has  $K + N$  variables,  $L$  linear inequality constraints, and  $L$  nonlinear inequality

constraints. Because the scaling  $(c\mathbf{w}^\natural, \frac{1}{c}\mathbf{x}^\natural)$  is consistent with the constraints for positive  $c$ , the program will return a solution where  $\|\mathbf{h}\|_2 = \|\mathbf{m}\|_2$ . Thus, if recovery is successful, the optimal solution is  $\left(\mathbf{h}^\natural \sqrt{\frac{\|\mathbf{m}^\natural\|_2}{\|\mathbf{h}^\natural\|_2}}, \mathbf{m}^\natural \sqrt{\frac{\|\mathbf{h}^\natural\|_2}{\|\mathbf{m}^\natural\|_2}}\right)$ .

## 1.2 Main Result

In this paper, we consider the bilinear recovery problem (1), where the subspaces given by  $\mathbf{B}$  and  $\mathbf{C}$  are random. Specifically, we show that if  $\mathbf{B}$  and  $\mathbf{C}$  have i.i.d. Gaussian entries, then exact recovery of  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is possible with nonzero probability when there are at least 4 times as many measurements as degrees of freedom.

**Theorem 1.** *Fix  $(\mathbf{h}^\natural, \mathbf{m}^\natural) \in \mathbb{R}^K \times \mathbb{R}^N$  such that  $\mathbf{h}^\natural \neq 0$  and  $\mathbf{m}^\natural \neq 0$ . Let  $\mathbf{B} \in \mathbb{R}^{L \times K}, \mathbf{C} \in \mathbb{R}^{L \times N}$  have i.i.d.  $\mathcal{N}(0, 1)$  entries. Then  $\left(\mathbf{h}^\natural \sqrt{\frac{\|\mathbf{m}^\natural\|_2}{\|\mathbf{h}^\natural\|_2}}, \mathbf{m}^\natural \sqrt{\frac{\|\mathbf{h}^\natural\|_2}{\|\mathbf{m}^\natural\|_2}}\right)$  is the unique solution to (BH) with probability at least*

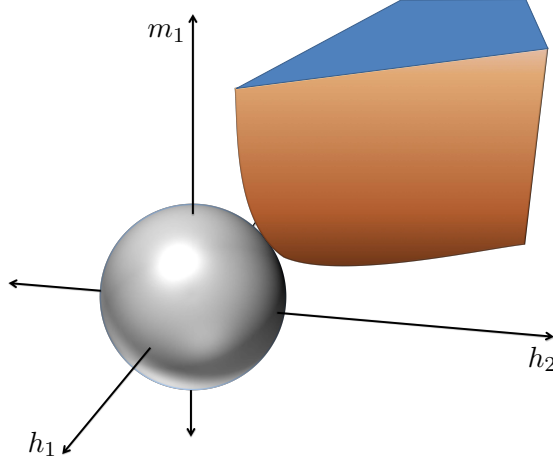
$$1 - \exp\left(-\frac{[L - (4N + 4K - 3)]^2}{2(L - 1)}\right),$$

*provided that  $L > 4N + 4K - 3$ .*

This theorem provides an explicit lower bound on the recovery probability by the convex program (BH). If  $L > 4N + 4K - 3$ , there is a nonzero probability of successful recovery. By taking  $L \geq \tilde{C}(N + K)$ , the probability of failure becomes at most  $e^{-\tilde{c}L}$ , for universal constants  $\tilde{C}$  and  $\tilde{c}$ . The scaling of  $L$  in terms of  $N + K$  is information theoretically optimal up to a constant factor. The proof of Theorem 1 follows from estimating the probability of covering a sphere by random hemispheres chosen from a nonuniform distribution.

## 1.3 Discussion

The BranchHull formulation is a novel convex relaxation for bilinear recovery from the entrywise product of vectors with known signs, and it enjoys a recovery guarantee when those vectors belong to random real subspaces of appropriate dimensions. The formulation is nothing more than finding which point of a convex set is closest to the origin. Geometrically, exact recovery is possible by  $\ell_2$ -norm minimization because the feasible set of  $(\mathbf{h}, \mathbf{m})$  has a ‘pointy’ ridge that corresponds to the fundamental scaling ambiguity, as illustrated in Figure 2.



**Figure 2:** An illustration of the geometry of BranchHull in the case where  $\mathbf{h} \in \mathbb{R}^2$  and  $\mathbf{m} \in \mathbb{R}^1$ . The feasible set of BranchHull has a shape similar to the solid in the top right. The ridge of this set corresponds to the fundamental scaling ambiguity of the bilinear recovery problem. The solution to BranchHull is given by the smallest scaling of the unit ball that intersects the feasible set. The minimizer is exactly on this ridge because the ridge is ‘pointy.’

An interesting feature of BranchHull is that it does not require an approximate solution or initialization in order to be solved or stated. In contrast, the Wirtinger gradient descent method for blind deconvolution uses an approximate solution as an initialization that ensures iterates begin in the basin of attraction of a global minimizer. Also in contrast, the PhaseMax linear program for phase retrieval uses an approximate solution, sometimes called an ‘anchor vector’, in its objective function. This anchor vector selects a vertex of a convex polytope that hopefully corresponds to the true solution. When this anchor vector aligns sufficiently with the true solution, then successful recovery is possible under information theoretically optimal sample complexity. Part of the role of the anchor vector is additionally to resolve the global phase ambiguity. In real bilinear recovery, there is also a corresponding global sign ambiguity that must be resolved. In our formulation,

the sign measurements resolve this ambiguity.

The idea of convex relaxations in the natural parameter space for bilinear problems is not new. For example, in nonlinear programming (NLP) or mixed integer nonlinear programming (MINLP) problems with bilinear constraints and specified variable bounds, a McCormick relaxation [18] replaces bilinear terms with four linear inequality constraints that define a convex quadrilateral that contains the hyperbola of feasible points within the variable bounds. Tighter convex relaxations are possible [6], such as by using the hyperbola itself as an inequality constraint [20]. These relaxations have been studied as part of branch and bound approaches to NLPs and MINLPs. Under certain conditions and branching rules [12] these approaches can find a global minimizer; however, the branching results in many convex programs that need to be solved, and it may result in exponential time complexity. In contrast, the present paper considers only the single convex program, BranchHull, achieved by the natural convex relaxation of bilinear constraints with only sign information. This work establishes conditions — in particular, subspace conditions — under which exact recovery by an efficient convex program can be rigorously established.

This work motivates several interesting and important extensions. Most immediately, BranchHull can be extended when the phases of complex vectors are known. Because of applications in signal processing and communications, it is also important to extend the theory for BranchHull to include deterministic subspaces, such as the span of partial Fourier matrices. Noise and outlier tolerance should be established for BranchHull or a variation with slack variables, such as in [10]. Outlier tolerance is particularly important because relatively few outlier measurements can substantially alter the shape of the feasible set. It would also be interesting to develop convex relaxations in the natural parameter space that do not use sign information. Further, extensions to more general bilinear recovery problems are of significant interest. All of these directions are left for future publications.

## 1.4 Notation

Vectors and matrices are written with boldface, while scalars and entries of vectors are written in plain font. For example,  $c_{\ell 1}$  is the first entry of  $\mathbf{c}_\ell$ . We write  $\mathbf{1}$  as the vector of all ones with dimensionality appropriate for the context. Let  $[L] = \{1, 2, \dots, L\}$ . Let  $\mathbf{e}_i$  be the  $i$ th standard basis element. We write  $K + N \lesssim L$  to mean that there exists a constant  $C$  such that  $K + N \leq CL$ . Given a vector in  $\mathbf{x} \in \mathbb{R}^N$ , let  $\tilde{\mathbf{x}} \in \mathbb{R}^{N-1}$  be the subvector formed by all but the first coefficient of  $\mathbf{x}$ . Let  $\mathbb{S}^{n-1}$  be the unit sphere in



$\mathbb{R}^n$ . For matrices  $\mathbf{A}, \mathbf{B}$ , let  $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{B}^\top \mathbf{A})$  be the Hilbert-Schmidt inner product of  $\mathbf{A}$  with  $\mathbf{B}$ . For a set  $S$ , let  $\text{Conv}(S)$  be its convex hull. Let  $\text{Re}\{z\}$  be the real part of a complex  $z$ .

## 2 Proof

To prove the theorem, we will show that  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is the unique minimizer of an optimization with a larger feasible set defined by linear constraints.

**Lemma 1.** *If  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is the unique solution to*

$$\begin{aligned} \underset{\mathbf{h} \in \mathbb{R}^K, \mathbf{m} \in \mathbb{R}^N}{\text{minimize}} \quad & \|\mathbf{h}\|_2^2 + \|\mathbf{m}\|_2^2 \text{ subject to } y_\ell \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \mathbf{h} \mathbf{m}^{\natural\top} + \mathbf{h}^\natural \mathbf{m}^\top \rangle \geq 2y_\ell^2, \\ & \ell = 1, \dots, L \end{aligned} \quad (2)$$

*then  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is the unique solution to (BH).*

*Proof of Lemma 1.* It suffices to show that the feasible set of (2) contains the feasible set of (BH). We may rewrite (BH) as

$$\begin{aligned} \underset{\mathbf{h} \in \mathbb{R}^K, \mathbf{m} \in \mathbb{R}^N}{\text{minimize}} \quad & \|\mathbf{h}\|_2^2 + \|\mathbf{m}\|_2^2 \text{ subject to } y_\ell \mathbf{b}_\ell^\top \mathbf{h} \cdot \mathbf{c}_\ell^\top \mathbf{m} \geq y_\ell^2 \\ & s_\ell \cdot \mathbf{b}_\ell^\top \mathbf{h} \geq 0, \ell = 1, \dots, L. \end{aligned}$$

We now use the fact that a convex set with a smooth boundary is contained in a halfspace defined by the tangent hyperplane at any point on the boundary of the set. Consider the point  $(w_\ell^\natural, x_\ell^\natural) \in \mathbb{R}^2$ , and observe that

$$\left\{ (w_\ell, x_\ell) \in \mathbb{R}^2 \left| \begin{array}{l} y_\ell w_\ell x_\ell \geq y_\ell^2 \\ \text{sign}(w_\ell) = s_\ell \end{array} \right. \right\} \subseteq \left\{ (w_\ell, x_\ell) \in \mathbb{R}^2 \left| \begin{pmatrix} y_\ell x_\ell^\natural \\ y_\ell w_\ell^\natural \end{pmatrix} \cdot \begin{pmatrix} w_\ell - w_\ell^\natural \\ x_\ell - x_\ell^\natural \end{pmatrix} \geq 0 \right. \right\}. \quad (3)$$

Plugging in  $w_\ell = \mathbf{b}_\ell^\top \mathbf{h}$  and  $x_\ell = \mathbf{c}_\ell^\top \mathbf{m}$ , we have that any feasible  $(\mathbf{h}, \mathbf{m})$  satisfies

$$y_\ell \mathbf{c}_\ell^\top \mathbf{m}^\natural \mathbf{b}_\ell^\top \mathbf{h} + y_\ell \mathbf{b}_\ell^\top \mathbf{h}^\natural \mathbf{c}_\ell^\top \mathbf{m} \geq 2y_\ell^2, \quad \ell = 1, \dots, L,$$

which implies  $y_\ell \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \mathbf{h} \mathbf{m}^{\natural\top} + \mathbf{h}^\natural \mathbf{m}^\top \rangle \geq 2y_\ell^2$  for all  $\ell$ .  $\square$

We now show that  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is the unique solution to the optimization problem (2) if the unit sphere in  $\mathbb{R}^{N+K-2}$  is covered by  $L$  hemispheres given in terms of  $\mathbf{b}_\ell$  and  $\mathbf{c}_\ell$ . Write  $\mathbf{b}_\ell = (b_{\ell 1}, \tilde{\mathbf{b}}_\ell)$ , where  $\tilde{\mathbf{b}}_\ell$  contains all but the first elements of  $\mathbf{b}_\ell$ . Similarly, write  $\mathbf{c}_\ell = (c_{\ell 1}, \tilde{\mathbf{c}}_\ell)$ .

**Lemma 2.** Let  $\mathbf{h}^\natural = \mathbf{e}_1$  and  $\mathbf{m}^\natural = \mathbf{e}_1$ . The unique solution to (2) is  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  if for all  $(\widetilde{\delta\mathbf{m}}, \widetilde{\delta\mathbf{h}}) \in \mathbb{R}^{N-1} \times \mathbb{R}^{K-1}$  there exists an  $\ell \in [L]$  such that  $b_{\ell 1} \neq 0$ ,  $c_{\ell 1} \neq 0$ , and

$$\left\langle \frac{\tilde{\mathbf{c}}_\ell}{c_{\ell 1}}, \widetilde{\delta\mathbf{m}} \right\rangle + \left\langle \frac{\tilde{\mathbf{b}}_\ell}{b_{\ell 1}}, \widetilde{\delta\mathbf{h}} \right\rangle \leq 0. \quad (4)$$

*Proof of Lemma 2.* Because the feasible set of (2) is closed and convex, and because a closed convex set has a unique point closest to the origin, (2) has a unique minimizer.

Consider a feasible point  $(\mathbf{h}^\natural + \delta\mathbf{h}, \mathbf{m}^\natural + \delta\mathbf{m})$ . To prove that  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is a minimizer of (2), it suffices to show

$$\langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \mathbf{h}^\natural \mathbf{m}^{\natural\top} \rangle \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \delta\mathbf{h} \mathbf{m}^{\natural\top} + \mathbf{h}^\natural \delta\mathbf{m}^\top \rangle \geq 0 \quad \forall \ell \Rightarrow \langle \mathbf{m}^\natural, \delta\mathbf{m} \rangle + \langle \mathbf{h}^\natural, \delta\mathbf{h} \rangle \geq 0.$$

Plugging in  $\mathbf{h}^\natural = \mathbf{e}_1$  and  $\mathbf{m}^\natural = \mathbf{e}_1$ , it suffices to show

$$b_{\ell 1} c_{\ell 1} [b_{\ell 1} c_{\ell 1} (\delta m_1 + \delta h_1) + b_{\ell 1} \tilde{\mathbf{c}}_\ell^\top \widetilde{\delta\mathbf{m}} + c_{\ell 1} \tilde{\mathbf{b}}_\ell^\top \widetilde{\delta\mathbf{h}}] \geq 0 \quad \forall \ell \Rightarrow \delta m_1 + \delta h_1 \geq 0$$

Dividing by  $b_{\ell 1}^2 c_{\ell 1}^2$ , it suffices to show

$$\begin{aligned} \delta m_1 + \delta h_1 + \left\langle \frac{\tilde{\mathbf{c}}_\ell}{c_{\ell 1}}, \widetilde{\delta\mathbf{m}} \right\rangle + \left\langle \frac{\tilde{\mathbf{b}}_\ell}{b_{\ell 1}}, \widetilde{\delta\mathbf{h}} \right\rangle &\geq 0 \quad \forall \ell \text{ s.t. } b_{\ell 1} \neq 0 \text{ and } c_{\ell 1} \neq 0 \\ &\Rightarrow \delta m_1 + \delta h_1 \geq 0. \end{aligned}$$

To prove this, it suffices to prove

$$\begin{aligned} \forall (\widetilde{\delta\mathbf{h}}, \widetilde{\delta\mathbf{m}}) \in \mathbb{R}^{N-1} \times \mathbb{R}^{K-1}, \exists \ell \text{ s.t. } b_{\ell 1} \neq 0, c_{\ell 1} \neq 0, \\ \text{and } \left\langle \frac{\tilde{\mathbf{c}}_\ell}{c_{\ell 1}}, \widetilde{\delta\mathbf{m}} \right\rangle + \left\langle \frac{\tilde{\mathbf{b}}_\ell}{b_{\ell 1}}, \widetilde{\delta\mathbf{h}} \right\rangle \leq 0. \end{aligned}$$

□

For a given vector  $\mathbf{a}$ , we will call  $\{\boldsymbol{\delta} \in \mathbb{S}^{n-1} : \langle \mathbf{a}, \boldsymbol{\delta} \rangle \geq 0\}$  the hemisphere centered at  $\mathbf{a}$ . We now provide a lower bound to the probability of covering the unit sphere by hemispheres centered at  $m$  random directions under a nonuniform probability distribution that is symmetric to negation. This lemma is an immediate generalization of Lemma 2 in [9], with a nearly identical proof.

**Lemma 3.** Choose  $m$  independent random vectors  $\{\mathbf{a}_i\}_{i=1}^m$  in  $\mathbb{S}^{n-1}$  from a (possibly nonuniform) distribution that is symmetric with respect to negation, and is such that all subsets of size  $n$  are linearly independent with probability 1. Then, the hemispheres centered at  $\{\mathbf{a}_i\}_{i=1}^m$  cover the whole sphere with probability

$$1 - \frac{1}{2^{m-1}} \sum_{k=0}^{n-1} \binom{m-1}{k}.$$

This value is the probability of flipping at least  $n$  heads among  $m-1$  tosses.

*Proof of Lemma 3.* Classical arguments in sphere covering [28] show<sup>2</sup> the following: If  $m$  hyperplanes containing the origin are such that the normal vectors to any subset of  $n$  hyperplanes are linearly independent, then the complement of the union of these hyperplanes is partitioned into

$$r(n, m) = 2 \sum_{k=0}^{n-1} \binom{m-1}{k}$$

connected regions. In each of these regions, every point lies on the same side of each hyperplane. Alternatively put, each region corresponds to a unique assignment of a side of each hyperplane. For a fixed set of  $m$  hyperplanes, if the half space on either side of each hyperplane is selected by independent tosses of a fair coin, then with probability at least that given in the lemma statement, there will be no nontrivial intersection all of these half spaces.

By the assumption that the distribution of  $\mathbf{a}_i$  is symmetric with respect to negation, we have that for any  $\mathbf{z}_i \in \mathbb{S}^{n-1}$ , the conditional distribution of  $\mathbf{a}_i$  given  $\mathbf{a}_i \in \{\pm \mathbf{z}\}$  is uniform over the two elements  $\pm \mathbf{z}$ . By independence, for any fixed  $\{\mathbf{z}_i\}_{i=1}^m \in (\mathbb{S}^{n-1})^m$ , the distribution of  $\{\mathbf{a}_i\}$  conditioned on the event  $\{\mathbf{a}_i = \pm \mathbf{z}_i\}$  is uniform over the  $2^m$  possibilities. Thus, conditioned on this event, the probability that the sphere is covered is at least that of the lemma statement. Integrating over all possible  $\{\mathbf{z}_i\}$ , the lemma follows.  $\square$

Our last technical lemma provides an explicit lower bound to the probability of the sphere covering given in Lemma 2.

**Lemma 4.** With probability at least  $1 - \exp\left(-\frac{[(L-1)-2(2N+2K-2)]^2}{2(L-1)}\right)$ , we have that

$$\forall (\widetilde{\delta \mathbf{h}}, \widetilde{\delta \mathbf{m}}) \in \mathbb{R}^{N-1} \times \mathbb{R}^{K-1}, \exists \ell \text{ such that } \left\langle \frac{\tilde{\mathbf{c}}_\ell}{c_{\ell 1}}, \widetilde{\delta \mathbf{m}} \right\rangle + \left\langle \frac{\tilde{\mathbf{b}}_\ell}{b_{\ell 1}}, \widetilde{\delta \mathbf{h}} \right\rangle \leq 0. \quad (5)$$

---

<sup>2</sup>This article credits [21] for the proof argument.

*Proof of Lemma 4.* To show (5), we must show that the  $L$  hemispheres (of the unit sphere in  $\mathbb{R}^{N-1} \times \mathbb{R}^{K-1}$ ) centered at  $(-\frac{\tilde{\mathbf{c}}_\ell}{c_{\ell 1}}, -\frac{\tilde{\mathbf{b}}_\ell}{b_{\ell 1}})$  cover the entire sphere. As the distribution of  $(\frac{\tilde{\mathbf{c}}_\ell}{c_{\ell 1}}, \frac{\tilde{\mathbf{b}}_\ell}{b_{\ell 1}})$  is invariant to negation, and as any  $n$  samples from this distribution are linearly independent with probability 1, Lemma 3 gives that the probability that (5) holds is at least the probability of flipping at least  $2N + 2K - 2$  heads among  $L - 1$  tosses of a fair coin.

We now bound the probability of getting at least  $n$  heads among  $m$  fair coin tosses. Let  $X$  be the number of heads in  $m$  tosses. By Hoeffding's inequality for Bernoulli random variables [27], for any  $t \geq 0$ ,

$$\mathbb{P}\left(X - \frac{m}{2} < -mt\right) \geq 1 - e^{-2mt^2}.$$

By selecting  $t = \frac{1}{2} - \frac{n}{m}$ , we get that when  $n \leq m/2$ ,

$$\mathbb{P}(\text{at least } n \text{ heads among } m \text{ tosses}) \geq 1 - e^{-2m(\frac{1}{2} - \frac{n}{m})^2} = 1 - e^{-\frac{(m-2n)^2}{2m}}.$$

The lemma follows by plugging in  $m = L - 1$  and  $n = 2N + 2K - 2$  into the above probability estimate.  $\square$

Now, we may prove the theorem.

*Proof of Theorem 1.* Without loss of generality, let  $\|\mathbf{h}^\natural\|_2 = \|\mathbf{m}^\natural\|_2$ . This is possible because for any  $\mathbf{b}_\ell$  and  $\mathbf{c}_\ell$ , we have that  $\left(\mathbf{h}^\natural \sqrt{\frac{\|\mathbf{m}^\natural\|_2}{\|\mathbf{h}^\natural\|_2}}, \mathbf{m}^\natural \sqrt{\frac{\|\mathbf{h}^\natural\|_2}{\|\mathbf{m}^\natural\|_2}}\right)$  and  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  give equal values of  $y_\ell = \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \mathbf{h}^\natural \mathbf{m}^{\natural \top} \rangle$ .

Further, without loss of generality, let  $\|\mathbf{h}^\natural\|_2 = \|\mathbf{m}^\natural\|_2 = 1$ . This is possible because the scaling

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{\|\mathbf{h}^\natural\|_2}, \quad \hat{\mathbf{m}} = \frac{\mathbf{m}}{\|\mathbf{m}^\natural\|_2}, \quad \hat{\mathbf{h}}^\natural = \frac{\mathbf{h}^\natural}{\|\mathbf{h}^\natural\|_2}, \quad \hat{\mathbf{m}}^\natural = \frac{\mathbf{m}^\natural}{\|\mathbf{m}^\natural\|_2},$$

turns (BH) into

$$\begin{aligned} & \underset{\mathbf{h} \in \mathbb{R}^K, \mathbf{m} \in \mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{h}^\natural\|_2^2 \|\hat{\mathbf{h}}\|_2^2 + \|\mathbf{m}^\natural\|_2^2 \|\hat{\mathbf{m}}\|_2^2 \\ & \text{subject to} \quad \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \hat{\mathbf{h}}^\natural \hat{\mathbf{m}}^{\natural \top} \rangle \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \hat{\mathbf{h}} \hat{\mathbf{m}}^\top \rangle \geq \langle \mathbf{b}_\ell \mathbf{c}_\ell^\top, \hat{\mathbf{h}}^\natural \hat{\mathbf{m}}^{\natural \top} \rangle^2 \\ & \quad s_\ell \cdot \mathbf{b}_\ell^\top \hat{\mathbf{h}} \geq 0, \quad \ell = 1, \dots, L. \end{aligned} \tag{6}$$

Further, without loss of generality we may take  $\mathbf{h}^\natural = \mathbf{e}_1$  and  $\mathbf{m}^\natural = \mathbf{e}_1$ . To see this is possible, let  $\mathbf{R}_{\mathbf{h}^\natural}$  and  $\mathbf{R}_{\mathbf{m}^\natural}$  be rotation matrices that map  $\mathbf{h}^\natural$  and  $\mathbf{m}^\natural$

to  $\mathbf{e}_1$ , respectively. Letting  $\bar{\mathbf{h}} = \mathbf{R}_{\mathbf{h}^\natural} \mathbf{h}$ ,  $\bar{\mathbf{m}} = \mathbf{R}_{\mathbf{m}^\natural} \mathbf{m}$ , and  $\bar{s}_\ell = \text{sign}(\mathbf{b}_\ell^\top \mathbf{R}_{\mathbf{h}^\natural}^\top \mathbf{e}_1)$ , problem (BH) can be written

$$\begin{aligned} & \underset{\mathbf{h} \in \mathbb{R}^K, \mathbf{m} \in \mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{R}_{\mathbf{h}^\natural} \bar{\mathbf{h}}\|_2^2 + \|\mathbf{R}_{\mathbf{m}^\natural} \bar{\mathbf{m}}\|_2^2 \\ & \text{subject to} \quad \langle \mathbf{R}_{\mathbf{h}^\natural} \mathbf{b}_\ell \mathbf{c}_\ell^\top \mathbf{R}_{\mathbf{m}^\natural}^\top, \mathbf{e}_1 \mathbf{e}_1^\top \rangle \langle \mathbf{R}_{\mathbf{h}^\natural} \mathbf{b}_\ell \mathbf{c}_\ell^\top \mathbf{R}_{\mathbf{m}^\natural}^\top, \bar{\mathbf{h}} \bar{\mathbf{m}}^\top \rangle \geq \langle \mathbf{R}_{\mathbf{h}^\natural} \mathbf{b}_\ell \mathbf{c}_\ell^\top \mathbf{R}_{\mathbf{m}^\natural}^\top, \mathbf{e}_1 \mathbf{e}_1^\top \rangle^2 \\ & \quad \bar{s}_\ell \cdot \mathbf{b}_\ell^\top \mathbf{R}_{\mathbf{h}^\natural}^\top \bar{\mathbf{h}} \geq 0, \quad \ell = 1, \dots, L. \end{aligned} \quad (7)$$

As  $\ell_2$  norms are invariant to rotation and as  $\mathbf{R}_{\mathbf{h}^\natural} \mathbf{b}_\ell$  and  $\mathbf{R}_{\mathbf{m}^\natural} \mathbf{c}_\ell$  have independent  $\mathcal{N}(0, 1)$  entries, we may take  $(\mathbf{h}^\natural, \mathbf{m}^\natural) = (\mathbf{e}_1, \mathbf{e}_1)$ .

Let  $E$  be the event that (5) holds. By Lemma 4,

$$\mathbb{P}(E) \geq 1 - \exp\left(-\frac{[(L-1) - 2(2N+2K-2)]^2}{2(L-1)}\right).$$

By Lemma 2, on  $E$ ,  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is the unique solution to (2). By Lemma 1, on  $E$ ,  $(\mathbf{h}^\natural, \mathbf{m}^\natural)$  is the unique solution to (BH). □

## Acknowledgements

PH acknowledges funding by the grant NSF DMS-1464525.

## References

- [1] Alireza Aghasi, Barmak Heshmat, Albert Redo-Sanchez, Justin Romberg, and Ramesh Raskar. Sweep distortion removal from terahertz images via blind demodulation. *Optica*, 3(7):754–762, 2016.
- [2] Ali Ahmed and Laurent Demanet. Leveraging diversity and sparsity in blind deconvolution. *arXiv preprint arXiv:1610.06098*, 2016.
- [3] Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Trans. Inform. Theory*, 60(3):1711–1732, 2014.
- [4] Sohail Bahmani and Justin Romberg. Phase retrieval meets statistical learning theory: A flexible convex relaxation. *arXiv preprint arXiv:1610.04210*, 2016.
- [5] Emmanuel Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Trans. Inform. Theory*, 61(4):1985–2007, 2015.

- [6] Pedro M Castro. Tightening piecewise mccormick relaxations for bilinear problems. *Computers & Chemical Engineering*, 72:300–311, 2015.
- [7] Yuxin Chen and Emmanuel Candes. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances Neural Inform. Process. Syst.*, pages 739–747, 2015.
- [8] James R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.
- [9] Tom Goldstein and Christoph Studer. Phasemax: Convex phase retrieval via basis pursuit. *arXiv preprint arXiv:1610.07531*, 2016.
- [10] Paul Hand and Vladislav Voroninski. Corruption robust phase retrieval via linear programming. *arXiv preprint arXiv:1612.03547*, 2016.
- [11] Paul Hand and Vladislav Voroninski. An elementary proof of convex phase retrieval in the natural parameter space via the linear program phasemax. *arXiv preprint arXiv:1611.03935*, 2016.
- [12] Reiner Horst and Hoang Tuy. *Global optimization: Deterministic approaches*. Springer Science & Business Media, 2013.
- [13] Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.
- [14] Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. *IEEE signal processing magazine*, 13(3):43–64, 1996.
- [15] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.
- [16] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *arXiv preprint arXiv:1606.04933*, 2016.
- [17] Shuyang Ling and Thomas Strohmer. Self-calibration and biconvex compressive sensing. *Inverse Problems*, 31(11):115002, 2015.
- [18] Garth P McCormick. Computability of global solutions to factorable nonconvex programs: Part i – convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976.

- [19] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. In *Advances Neural Inform. Process. Syst.*, pages 2796–2804, 2013.
- [20] Ignacio Quesada and Ignacio E Grossmann. A global optimization algorithm for linear fractional and bilinear programs. *Journal of Global Optimization*, 6(1):39–76, 1995.
- [21] Ludwig Schläfli. *Gesammelte mathematische Abhandlungen: 1814-1895*. 2. Birkhäuser, 1953.
- [22] Thomas G Stockham, Thomas M Cannon, and Robert B Ingebreetsen. Blind deconvolution through digital signal processing. *Proceedings of the IEEE*, 63(4):678–692, 1975.
- [23] Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 2379–2383. IEEE, 2016.
- [24] Ivana Todic and Pascal Frossard. Dictionary learning. *IEEE Signal Processing Magazine*, 28(2):27–38, 2011.
- [25] Stephen Tu, Ross Boczar, Max Simchowitz, Mahdi Soltanolkotabi, and Benjamin Recht. Low-rank solutions of linear matrix equations via procrustes flow. *arXiv preprint arXiv:1507.03566*, 2015.
- [26] Gang Wang, Georgios B Giannakis, and Yonina C Eldar. Solving systems of random quadratic equations via truncated amplitude flow. *arXiv preprint arXiv:1605.08285*, 2016.
- [27] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.
- [28] James G Wendel. A problem in geometric probability. *Math. Scand*, 11:109–111, 1962.